Data mining to improve process and energy performance

Knowledge Process Software plc

ICI Thornton Power Station

Carlsberg Tetley Ltd

Blackburn Starling Ltd

- Identify savings
- Payback periods < 1 year
- Turn data into knowledge
- Identify new opportunities
- Quantify benefits
- Sustain performance
- Highly cost effective
- Widely applicable

ENERGY EFFICIENCY

BEST PRACTICE PROGRAMME

FUTURE PRACTICE PROFILE 67

PARTICIPATING COMPANIES

The aim of this project was to demonstrate the value of the data typically collected by modern process monitoring and control systems, and specifically to show how this data could be used to identify opportunities to improve energy efficiency.

Data mining techniques were used to analyse historical process operating data, and have proved effective and easy to use. A methodology for the application of these techniques in the process industries has been developed, along with a supporting software toolkit.

Knowledge-based systems have been applied to analyse process data as it becomes available, providing timely advice to operators and engineers about process conditions and faults. Optimisers have been developed for refrigeration systems.

The project has stimulated much interest, and a number of leading process companies are making use of the results. We expect that the techniques will be taken up widely.

David Hart
Chief Operating Officer
Knowledge Process Software

David Gould
Operations Director
Blackburn Starling & Co Ltd

Akeel Al-Attar
Chief Technical Officer
Knowledge Process Software

Development work was carried out by Knowledge Process Software
Tel: 0870 6087202, www.knowledgeprocesssoftware.com,
a company formed from a collaboration between Hart Consultants and Attar Software Ltd.
Support was provided by Blackburn Starling & Co Ltd (Tel: 01159 866331).
ICI Thornton Power Station, Carlsberg Tetley Alloa Brewery, and
Bass Burton Brewery (formerly Carlsberg Tetley Burton Brewery) acted as host sites for the project

There may be other suppliers of similar systems in the market.
Please consult your supply directories or contact Environment and Energy Helpline 0800 585794
who may be able to provide more details on request.

FPP 67 First published 03/02

INTRODUCTION

BACKGROUND

Modern process monitoring and control systems can be used to improve process and energy performance. They provide operators and engineers with comprehensive information about process operations, they provide a flexible infrastructure for conventional and advanced control, and, importantly in the context of this project, they have the ability to acquire and store large amounts of process operating data, making this data readily available in a timely manner.

Process operating data can be highly valuable. It represents how plant and processes operate in practice, and 'hidden' within it, is useful knowledge, about:

- energy efficiency;
- emissions;
- yield;
- quality;
- throughput;
- reliability.

Finding this knowledge can be difficult, but the benefits can be huge; improved understanding of performance variations and the identification and quantification of new opportunities for improvement.

Data mining refers to a set of techniques that can be used to effectively analyse data - turning the data overload into knowledge. In this project, these techniques have been applied to illustrate the potential to find patterns, which relate to energy and process performance.

Data mining techniques are used widely, but their application to process and energy data in this project was innovative. They have a number of unique capabilities; the ability to handle massive data volumes (many thousands or even millions of records), to cope with situations where many factors are affecting performance, to discover patterns automatically and to express these patterns in an understandable form (as rules) thereby improving insight into operations.

Knowledge-based systems provide a further opportunity to improve performance. These systems can be used to 'capture' expertise, and then apply that expertise consistently, accurately and quickly. They can be used to interpret process operating data as it becomes available, and provide advice to operators and engineers, and even to implement actions automatically.

Knowledge-based systems (expert systems, rule-based systems) are not new. The expertise is generally in the form of procedures to perform calculations and rules, and the generation and maintenance of a consistent and comprehensive set of rules. This is generally a time consuming and costly task. In this project, techniques to generate rules quickly and effectively have been demonstrated.

OBJECTIVES

There were three main objectives of the project:

- to demonstrate the potential of data mining and knowledge-based systems to improve process and energy efficiency;
- to develop methodologies for the use of these techniques;
- to develop supporting software toolkits for data mining and knowledge-based systems.

Carlsberg Tetley brewhouse

USES

There are many potential applications of data mining and knowledge-based systems.

Data mining techniques can be used by almost any operating company where operating data is collected, and the techniques used to address the many aspects of process performance. This Profile looks at the use of data mining to address refrigeration efficiency and power station operations. Other projects initiated directly as a result of this project include ones which address:

- potash refinery yield;
- rotary dryer energy use;
- toner manufacturing performance - yield and energy;
- kiln operations;
- oil refinery processing;
- oil field operations;
- corrosion rates;
- paper manufacturing.

Knowledge-based systems likewise have wide application, to provide timely advice on process operations to engineers and operators. Their use to improve refrigeration performance is discussed below.

DATA MINING

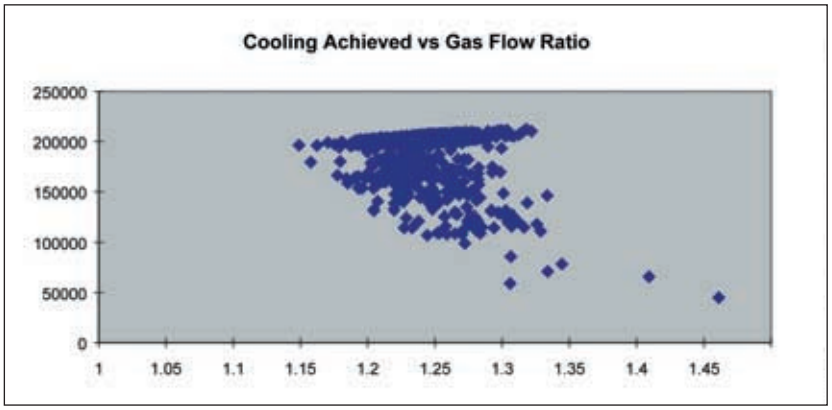

Fig 1 X-Y plot of cooling achieved versus gas flow in heat exchangers

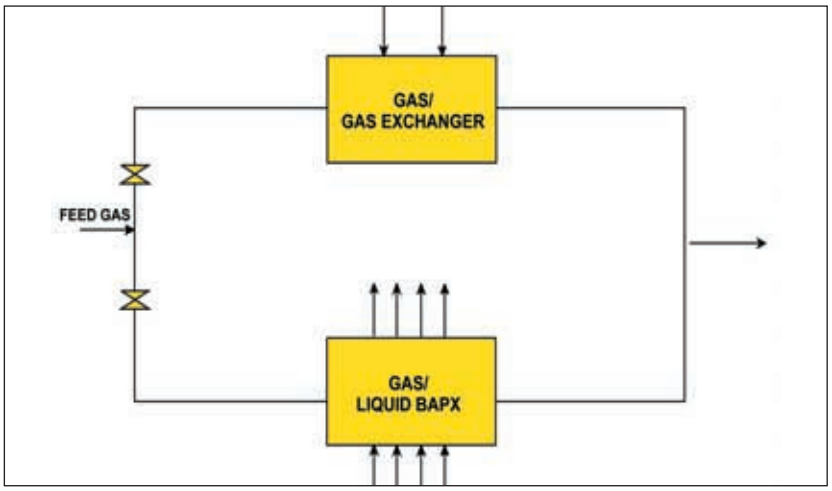

Fig 2 Parallel heat exchangers

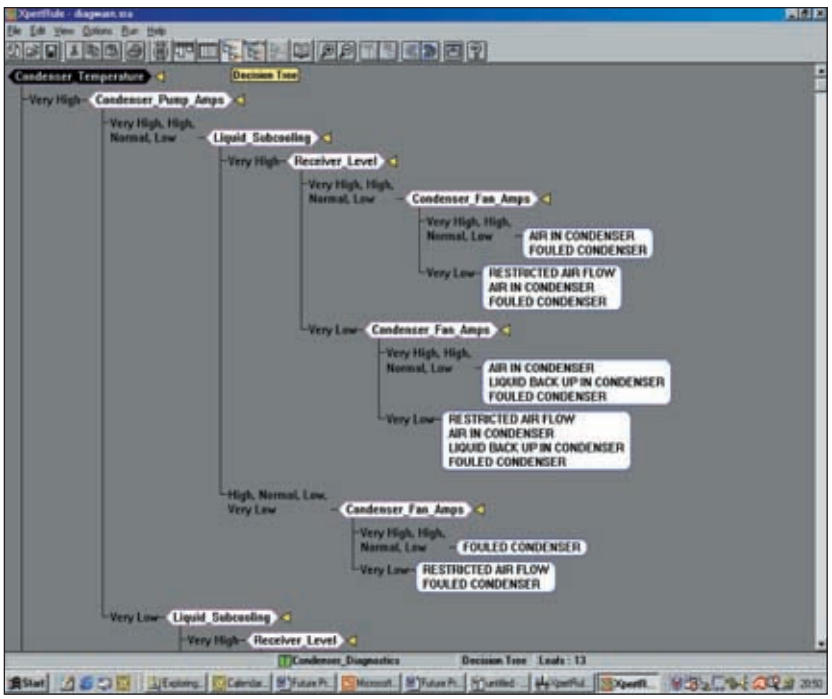

Fig 3 Decision tree - heat exchanger performance, showing average cooling achieved

WHAT IS DATA MINING?

Data mining refers to a set of techniques to make better use of data. They are widely used, for example to analyse financial, insurance and marketing data. In this project, the techniques have been used to find patterns in process operating data, the data that would typically be collected by modern process monitoring and control systems such as DCS and SCADA systems.

Finding knowledge in data can be a difficult task, especially where data volumes are huge and where the relationships between variables are complex. Traditional techniques (database queries and graphical analysis by spreadsheet) have limitations.

In Fig 1, for example, the cooling achieved by process heat exchangers operating in parallel is plotted against the ratio of the flows through the two exchangers. The heat exchangers are shown in Fig 2. It is difficult to identify whether a high flow ratio is better than a low ratio, but clearly such knowledge would be useful and simple to implement.

The main technique employed in data mining is rule induction. With this technique, patterns in data are expressed as a set of rules, which are often displayed graphically in the form of a decision tree. Part of a tree, relating to the parallel heat exchanger problem, is shown in Fig 3.

The tree shows that the main influence on the cooling achieved, is the overall feed rate to the two exchangers. It also shows that at high feed rates, a higher flow ratio increases the cooling achieved. To improve performance the operators simply need to operate with higher flow ratios at the higher feed rates, an improvement, which requires no investment to implement.

The above is a simple example to illustrate how data mining works. The techniques can find more subtle patterns, the main, special characteristics of the techniques are:

- The ability to handle massive databases with many attributes. It is feasible to handle hundreds of thousands of records (e.g. data collected every minute from a process for a year or more) and hundreds of parameters.
- The patterns are found automatically.
- The patterns are expressed in an understandable form, which provides the analyst/engineer with insight into operations.

A feature of a data mining analysis is that the opportunities discovered will most likely involve simple changes to control and operating practices, requiring little or no investment to implement.

DATA MINING APPLICATION STUDIES

DATA MINING METHODOLOGY

The stages of a data mining study are as follows:

- *Examine* the process/utility system to gain an understanding of current operations, performance and issues, and the available data.
- *Define* the objectives of the data analysis (typically termed outcomes). These might be energy use, yield, production rate and similar, or perhaps a combination of these objectives expressed in a common currency, for example operating cost.
- *Identify* the possible influencing factors (commonly termed attributes). These are parameters which could feasibly affect the outcomes, and may be raw measured values, or calculated values (for example a flow ratio), which are termed derived attributes. The attributes may be values measured at the same time as the outcome, earlier values to take account of time delays, or they might be a calculated value, which summarises events leading up to the time of the outcome.
- *Define* the data requirements, including the amount and frequency required.
- *Acquire* the data, and then pre-process the data. Pre-processing includes checking the data and calculating the derived attributes, to produce a data mining database. This database may also be compressed to allow faster processing in memory. The data will typically be divided into training and test data-sets so that any patterns found in training data may be validated against unseen test data.
- *Mine* to identify the patterns, followed by interpretation of the patterns to identify and quantify opportunities for improvement.
- *Implement* and then *monitor* to check that improvements have been achieved. Repeat the data mining stages to ensure that performance is sustained and to identify new opportunities.

The cost of an initial data mining study will depend on the complexity of the process being analysed, and might typically be £10,000 to £50,000, and achieve a three month payback.

REFRIGERATION

Data from the Bass Burton Brewery (formerly Carlsberg Tetley Burton Brewery) Refrigeration System was mined to improve understanding of the variations in energy efficiency and to identify opportunities to reduce operating costs.

The system cools brine typically to -8°C which is pumped to the various process cooling heat exchangers and vessels within the brewery. There are five screw compressor packages, each with a nominal 1,050 kW cooling capacity. The evaporators are plate type, and the condensers evaporative units. The plant uses electricity worth some

Bass Burton Brewery refrigeration system

£350,000 a year at 1999 prices. The system was designed for high-efficiency and is the subject of Good Practice Case Study 248 available from the Energy Efficiency Best Practice Programme.

The outcome (or objective) of the data mining analysis was the 'Coefficient of Performance', or COP. This is defined as the cooling achieved by the plant divided by the power used. A higher value implies higher efficiency, and values of around 4 are typical for such a system. Factors likely to have an impact on COP (the attributes) are:

- ambient temperature;
- brine temperature;
- compressor selection;
- compressor loading.

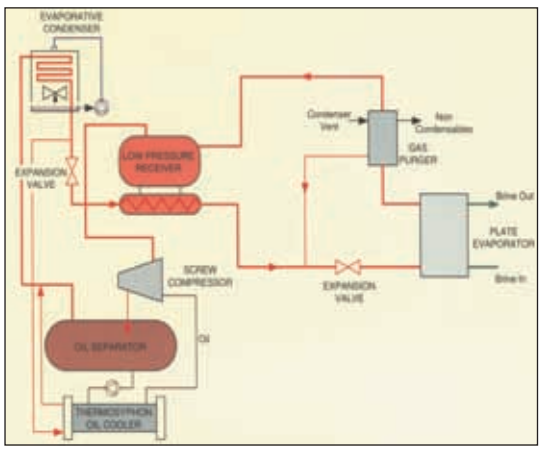

Fig 4 Diagram of single refrigeration package

DATA MINING APPLICATION STUDIES

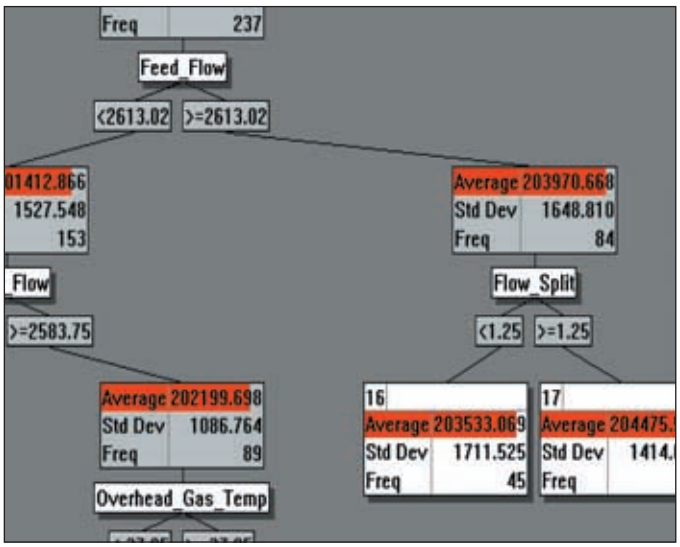

Fig 5 Part of a decision tree - refrigeration performance showing average COP

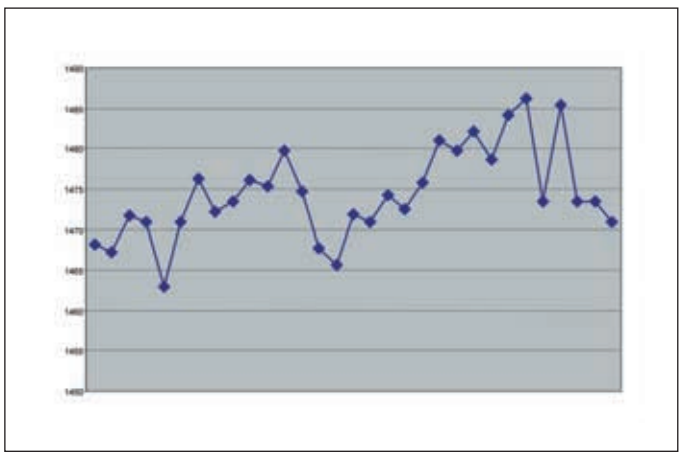

Fig 6 Manifold pressure variations over half an hour

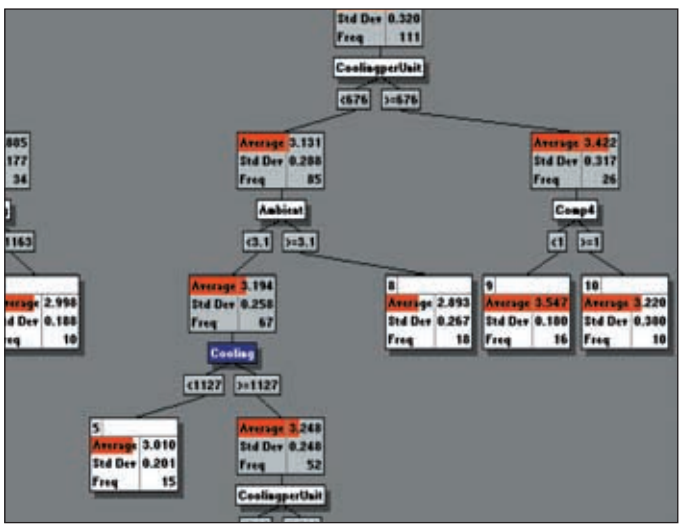

Fig 7 Part of a decision tree - power station performance showing average cost per unit of steam

DATA MINING APPLICATION STUDIES

A database of these attributes and the outcome, collected at a fifteen minute frequency, was mined. Part of a resulting decision tree is shown in Figure 5.

The analysis revealed two significant patterns:

- COP increases with increasing loading (Cooling PerUnit) on the screw compressors. Operation with compressors fully loaded reduces energy use considerably, by some 20%.
- COP decreases when compressor 4 is operating. All five compressors are of identical design, yet compressor 4 operates less efficiently.

The first of these patterns could be identified from first principles modelling. The data mining approach has several advantages, however:

- the patterns reflect what actually happens on the plant (and is not based on assumptions about compressor and heat exchanger performance which may well not be valid);
- the patterns were identified quickly and cheaply using readily available operating data. In contrast, first principles modelling would be costly and complex, and require a deep understanding of refrigeration principles. Furthermore, compressor performance data may not be readily available;
- the analysis not only reveals the importance of compressor loading on COP, but also the extent of the problem (poor conditions are encountered 75% of the time). This helped to reveal a control strategy problem that was unknown to the site at the time.

In total, energy savings of 12% were identified, (worth £42,000 a year), most of which could be achieved with little or no extra investment.

POWER STATION PERFORMANCE

An analysis of the efficiency of the ICI Thornton Power Station was also undertaken. The station housed three boilers, two of which were normally in service at any one time. The boilers were capable of firing natural gas or oil, and generated steam at up to 1,600 psig.

Electricity for the adjacent site was generated in three alternators, the primary turbo alternator taking steam at a nominal 1,500 psig and reducing it to 200 psig, and two secondary condensing units.

Data was collected from the power station after modification of the site monitoring systems, and mined, with the power station operating cost per unit of steam being generated as the outcome. Attributes included the boiler selection, boiler loadings, pressures and temperatures, bleed steam flows from the turbines and similar. Fig 6 shows the manifold steam pressure (boiler steam pressure) over a half-hour period.

Part of a decision tree is shown in Fig 7, which shows the impact of the manifold pressure (mpress) on the operating costs; a higher pressure reduces the operating cost per unit of steam generated.

By way of contrast, Fig 9 shows the cost plotted against the manifold pressure. The influence of manifold pressure is difficult to see, since many other factors which affect performance are also changing within the data set.

In total, energy savings of 4% were identified, (worth £250,000 a year). Some investment was required, but the payback period was less than 1 year.

KNOWLEDGE BASED SYSTEMS AND OPTIMISATION

KNOWLEDGE-BASED SYSTEMS

Knowledge-based systems (KBS - often referred to as rule-based systems or expert systems) are computer programs that contain expertise, usually in the form of rules and supporting procedures. They are used to apply expertise consistently, accurately and quickly. On utility and process systems their strength is their ability to identify operating problems and opportunities quickly (in real time) and to provide good advice (or take the appropriate actions themselves). The systems can be used to:

- provide timely advice to operators, engineers and managers;
- diagnose faults with utility and process plant;
- provide rule-based (fuzzy) control of complex plant to optimise performance.

This project used KBS tools with innovative features, including the ability to automatically derive rules from lists of examples and exceptions (see below), and the ability to incorporate rules learnt from operating data.

Historically, one of the major difficulties with KBS development has been the time required to generate a consistent set of rules. With the rule induction techniques applied in this project, this task is virtually automated, allowing KBS systems to be developed and maintained cheaply and quickly, making them more cost-effective and more likely to be widely taken up.

A KBS tool was developed to diagnose faults with refrigeration systems. This system accepts data from the refrigeration monitoring and control system and interprets that data by applying rules. Part of the rule set for diagnosing condenser faults is shown in Fig 8. These rules were induced from a list of the possible faults and their symptoms. The KBS tool provides operators with an early warning of a fault, such as a blocked nozzle, air in the condenser, fouling and other similar faults, and provides relevant advice on how to rectify matters.

OPTIMISATION

The benefits of optimising the operation of process and utility systems are well known. However, finding an optimum solution can be a complex and expensive task. For this reason, the use of optimisers is generally restricted to larger process operators; oil exploration, refining and petrochemicals. Even within these sectors, optimisers are quite rare.

Genetic algorithms offer an opportunity to implement optimisation more widely. The technique involves creating a 'first generation' of solutions to the problem where the variables are selected at random. These solutions are then assessed and the next generation of solutions is found by 'combining' the better solutions and discarding the rest - a process analogous to 'natural selection' - hence the term genetic algorithms. The process is computationally intensive but not beyond the power of modern PCs.

Complex, non-linear, constrained problems, where there are both continuous and discrete inputs, can be handled.

This project has also demonstrated the use of genetic algorithms. A refrigeration optimiser was developed which can select the best combinations of compressors, compressor load levels, chilled brine flow rates and brine temperatures, to meet the cooling demands of a brewery. The optimiser can take into account the future cooling demands and modify operations based on the half-hourly electricity prices (maximum demand management).

Optimisers can be used effectively as part of a data mining initiative to balance the various typical competing objectives, for example the need to improve quality, but at the same time increase yield and reduce energy cost.

The prototype optimisers were tested to demonstrate their ability to find the most efficient operating conditions, but were not fully deployed as part of this project. They are currently available as commercial products.

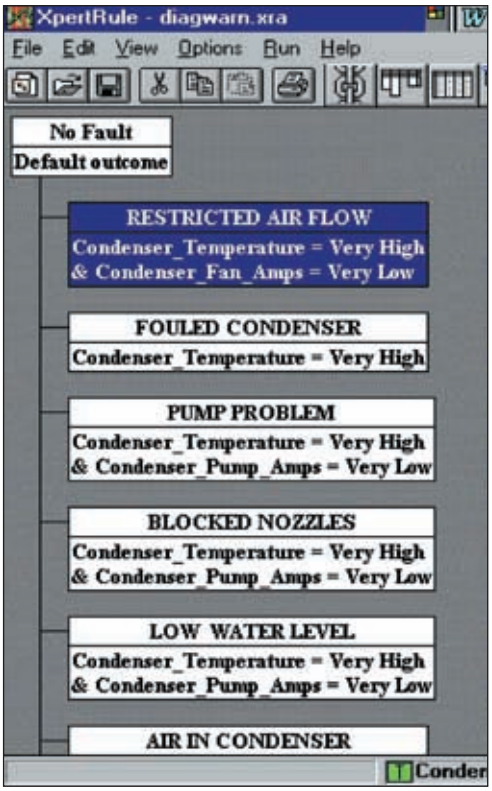

Fig 8 List of faults turned automatically into fault diagnosis rules

PROCESS OPTIMISATION TOOLKIT AND CONCLUSIONS

PROCESS OPTIMISATION TOOLKIT

An objective of this project was to develop software tools to allow the wider uptake of data mining and KBS technology. These technologies have the potential to make a significant impact on the use of energy in the process industries, and on process and energy performance more generally.

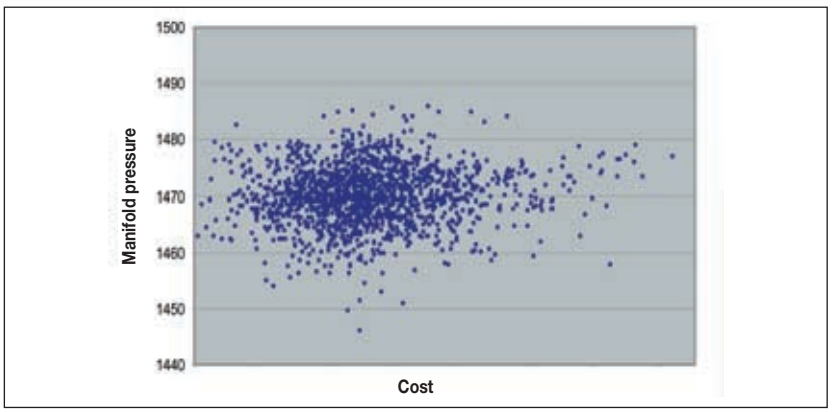

Fig 9 Power station manifold pressure vs operating cost

A suite of prototype tools has been developed during the project and as a result of additional contracts. The tools comprise:

- a data mining toolkit tailored to the process industries, which supports all aspects of an initial data mining study, including data pre-processing and mining;
- a data mining deployment tool, which is used to 'capture' a data mining analysis so that it can be re-used when new data becomes available;
- an expert system development toolkit, which can be used to develop and implement knowledge based systems, which integrates with process monitoring and control systems;
- an on-line performance monitoring tool, making use of models of process performance built from data and comparing the predicted use with actual values on line;
- an on-line fault diagnosis/advisory system;
- rule based control;
- process optimisation.

CONCLUSIONS

The project achieved the objectives of demonstrating the effectiveness of data mining, knowledge based systems and optimisers to identify and quantify opportunities to save energy and to sustain energy performance. A toolkit has been developed with a supporting methodology to facilitate wider uptake.

Knowledge Process Software has subsequently used the techniques developed in this project to identify energy savings at a number of large process sites in the UK and abroad. The company has now obtained substantial private finance to commercialise and market the technologies devloped in this project.

ICI Thornton Power Station